Filtering imaging records using wavelets and the Zakai equation

Ziad S. Haddad
Jet Propulsion laboratory
California Institute of Technology
4800 Oak Grove Boulevard
Pasadena, CA 91100
email: zsh@AlbertoVO5.jpl.nasa.gov

&

Santiago R. Simanca
Department of Mathematics
State University of New York
Stony Brook, New York 11794
email: santiago@math.sunysb.edu

4. “

Abstract

Consider the problem of detecting and localizing a front object moving in an "essentially stationary" backg round, using a sequence of two-dimensional low-SNR images of the scene. A natural approa ch consists of "dig itizing" each snapshot into a discrete set of observations, sufficiently (perhaps not exactly) matched to the object in question, then *tracking* the object using an appropriate stochastic filter. The tracking would be expected to make up for the low signal-tc-noise ratio, thus allowing one to "coherentl y" process successive images in order to beat down the noise and localize the object. Thus, "tracking" here does not refer to the ususal notion of detecting then tracking: rather, we track *in order to* detect. The problem then becomes one of choosing the appropriate image representation as well as the optimal (and necessarily non-linear) filter. We propose exact and approximate solutions using wavelets and the Zakai equation. The smoothness of the wavelets use d is required in the derivation of the evolution equation for the conditional density giving the filter, and their orthogonality makes it possible to carry out actual computations of the Ito -and change-of-gauge- terms in the algorithm effectively.

Acknowledgement

The first, author's work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract, with the National Aeronautics and Space Administration.

1 Introduction

Consider the problem of determining the position of an object moving on a line, in a plane, or in 3-space, using a sequence of snapshots of the "mostly stationary" region of space in which it evolves. Specifically, we would like to teach a computer to extract from the sequence of images the coordinates of the object of interest, in the case where the signal-to-noise ratio is very low. A natural first attempt to solve this problem consists of processing one image at a time, and looking for the pattern(s) we expect the object to produce on each frame individually. One would thus build templates of these patterns and move the templates all over the individual image, looking for the best "match". his first attempt is indeed firmly justified mathematically and would give a classical matched filter algorithm. Matched-filter processing is by now well-understood, and sophisticated considerations involving noise models, noise estimates, and various error probabilities allow the design of suitable matched-filter detectors. Thus, cases where good localization is possible on a single image can be considered solved, and shall not concern us further, except inasmuch as a tracking algorithm might make the detection-and-localization process on later images substantially more efficient. Rather, let us look at the problem where localization of the object on a single image is impossible or at least hard and fraught with ambiguity. For example, the object may not be "visible" at a non any image. Of course, if the object were moving in a straight line, one could add the successive images properly lined up along various possible bearings, thus hoping that the "signals" on the various images would add up "coherently" while all the rest, the "noise", would do so "incoherently": this would be the equivalent of the well-known delay-and-sum beamforming idea for plane waves in various settings. It would allow one to increase the "signal-to-noise" ratio sufficiently, given enough images (the number Of images would correspond, in this analogy, to the number of array elements), to bring the "spike" well above the noise threshold and thus make the required detection. Unfortunately, the time scales involved may make it highly unlikely that the moving object will have kept to a straight line track throughout the observation process; this modified matched-filter approach would thus fail in the presence of a typically "(maneuvering" object. in other cases, many parts of an image may be likely to provide equally good matches with the template, thus creating difficult ambiguity problems.

This filtering problem bears a definite resemblance to the typical problem solved by the Kalman-Bucy filter. Indeed, in both cases we are given asct of observations at a sequence of times (the sequence of images). Then, knowing how these observations arc affected by the coordinates of our object (i.e. knowing the pattern that the presence of the object creates on an individual image), and knowing the laws that must be obeyed by the motion of the object (in our case, to keep things simple, some continuity requirements and bounds on the velocity), we want to design an algorithm that will estimate as best can be the position of the object at every point in time. Tile algorithm can, and indeed should, make use of all past information in order to refine at each time step its best estimate for the new position of the object. Thus, this "tracking" would be expected to make up for the inevitable shortcomings of any attempted noise-reduction/background-suppression/signal-enhancement procedure applied to each individual image. One would then he able to find and track the object even if it were impossible to isolate on any single frame. The Kalman filler is just such an algorithm, except that it applies only to the linear case, where the ef-

feet of the object on the observations is a linear function of the coordinates of the object. In the case of images, that is of course never the case: the effect is by definition local, therefore the mathematical functions in question turn out to be at the very least bounded, and in any case far from linear.

Fortunately, a generalization of sorts of the Kalman filter to the non-linear case dots already exist, thanks to the work of T. Duncan, R. Mortensen, and M. Zakai. They derived the equations that must be solved in order to find the "optimal" filter (in the same least-squares sense as the Kalman case) which, when fed a set, of not,-necessarily-linear observations, will produce the best estimate of the required coordinates. The rest of this paper describes how their remarkable results, which until now have been mostly of a theoretical interest, can help solve our problem practically,

The most interesting application of our approach is to the case where the SNR is so low that the object is not dctcc.table on any single image. The idea of using stochastic filtering techniques to track objects on sequences of images is not, new (see e.g. [HN], [LY]). The stochastic approaches proposed to date typically use extended Kalman filters to estimate the motion parameters and track the object. Our purpose is different. Indeed, we do not assume that the object has been detected or localized, and we seek a way to combine a sufficient number of successive images (the exact number being in all likelihood inversely proportional to the signal-to-noise ratio) in order to detect the object. The extended Kalman filtering approach is unfortunately not appropriate for this problem because the dependence of the data (the images) on the variables to be estimated (the coordinates of the object) is not only non-linear, it is a non-analytic function (in fact, it is highly localized): using a Taylor series approximation, as the extended Kalman filtering approach requires, would therefore be totally inappropriate. Once the object is detected and roughly localized, however, tracking methods in the conventional sense (such as the ones derived from the approaches described in [HN] and [LY] for example) would then be quite applicable, and substantially more con iputationally efficient.

2 Mathematical formulation of the problem

To make the problem mathematically precise, let us assume that each image provides us concretely with a function on a compact region of $\mathbb{R}^N$ (with N = 1,2, or 3), representing the intensity of the image at every point in the region of interest. The problem will then be t we-fold: finding a good way to represent this sequence of functions, then finding a good way of extracting the position of the particle in question from the chronological sequence of pictures. The precise meaning of "good" in the image representation problem obviously depends on the method one chooses to track the particle. A natural choice is to look for a (necessarily nonlinear) filter. Viewing the position x_t of the object at time t as a stochastic time-varying vector, one will then have to estimate quantities which are functions of the stochastic process $\{x_t\}$ (for example its mean, its variance, etc.), on the basis of information obtained from related "directly observable' processes, namely the information in the sequence of images of the scene. Thus, one needs to identify these observation processes precisely, and to make explicit, their dependence on the object-position process x_t.

To that end, let us assume for simplicity, that the images are one-dimensional. One reasonable choice for the observation processes is to take O_t^i = the intensity of the image at time t over the subinterval $[i\delta, (i+1)\delta]$, where δ is

a positive real number chosen and fixed a priori, and r' runs over all integers. Symbolically, $O_t^i = \int_{i\delta}^{(i+1)\delta} s_t(x)dx$, where the function s_t represents the intensity of t he image at time t. We would then end up with observation processes O^i, each representing the intensity of the image over the pixel of width δ centered around (r' +1 /2)δ. This reasonable choice has one important drawback: in the language of [SM], it presupposes hat the images contain no detail structure finer than an a priori chosen scale, in the notation here δ. To get around this problem, let us rewrite the formula for each observation process as

$$O_t^i = \int_{J_{i\delta}}^{(i+1)\delta} s_t(x)dx = \int_{-\infty}^{\infty} \phi(\delta^{-1}x - i)s_t(x)dx, \quad (1)$$

where ϕ is the square window function $\phi(x) =:$ 1 if O < x < 1, and O otherwise. This new notation suggests that a natural way to replace the O_t^i's by processes which do not a priori eliminate finer details in the images (see [SM]) is to choose as our new processes

$$Z_t^{i,j} = \int_{-\infty}^{\infty} \psi_{i,j}(x)s_t(x)dx, \quad (2)$$

where $\psi_{0,0}$ is the Haar wavelet $\psi_{0,0}(x) = 1$ if $0 < x < 1/2$, --1 if $1/2 < x < 1$, and 0 otherwise, and $\psi_{i,j}(x) = 2^{\frac{j}{2}}\psi_{0,0}(2^j x - i)$. Here, i and j range over all the integers. At this stage, there is no need to restrict ourselves to the II aar wavelet. So let us assume that t he observat ion processes are chosen according to formula (2) but with $\psi_{0,0}$ any wavelet for which the family $\{\psi_{i,j}\}$ is an orthonormal basis for the space of all square-integrable functions (see, e.g., [ID1], [ID2], [SM]).

How do these observation processes depend on the object-position process x_t? Assume, for simplicity, that the object is "symmetric" in such a way that the intensity pattern it makes when it is placed, in any orientation, at the origin is always given by the function $r(x)$. In the l-dimensional case, this means that r is an even function (in t he 2-dimensional case, this would man that r is independent of polar angle). One can assume that the image at time t is then given by

$$s_t(x) = r(x - x_t) + (b_t(x) - \tilde{b}(x)).$$

The term in parentheses represents the background intensity b_t at time t, from which a template background intensity $\tilde{b}$ has been subtracted. The original assumption that the background is '(mostly stationary" was meant to imply that there is a given background $\tilde{b}$ from which b_t cannot differ much. Let us now make this more precise by requiring that the term in parentheses be the white noise process $N_t(x)$, and that $N_t(x), N_{t'}(x')$ be independent for all x different from x', t different from t'. 'I'hen, under these assumptions, the observations are given by

$$Z_t^{i,j} = \int_{-\infty}^{\infty} \psi_{i,j}(x)r(x - x_t)dx + \int_{-\infty}^{\infty} \psi_{i,j}(x)N_t(x)dx$$
$$= f_{i,j}(x_t) + N_t^{i,j},$$

where the functions $f_{i,j}$ are given by

$$f_{i,j}(y) = \int_{-\infty}^{\infty} \psi_{i,j}(x)r(x - y)dx,$$

while $N_t^{i,j}$ is similarly given by

$$N_t^{i,j} = \int_{-\infty}^{\infty} \psi_{i,j}(x)N_t(x)dx.$$

The assu mpt ion that $\{\psi_{i,j}\}$ is an orthonormal basis implies that each $N_t^{i,j}$ is again a white noise process and that $N_t^{i,j}, N_t^{i',j'}$ are independent if $(i,j) \# (i',j')$. In less precise terms, the orthonormal basis allowed us to represent the image sequence as a discrete set of observation processes in which the noise is no more correlated than it was in the original images. In the

more precise language of stochastic processes, each $N_t^{i,j}$ is written as $N_t^{i,j} dt = \sigma^{i,j} db_t^{i,j}$, where $b_t^{i,j}$ is standard Brownian motion, and where we will assume for simplicity that all the noise variances $\sigma^{i,j}$ are independent of (z, j), writing $\sigma^{i,j} = \sigma$ (this is equivalent to assuming that the distribution of the error incurred in subtracting $\hat{b}(x)$ from $b_t(x)$ is independent of the specific location x).

Our problem is now to look for a good estimate of the coordinates of the object x_t, given the observations $Z_t^{i,j}$ as in equation (2). Defining $y_t^{i,j}$ to be $y_t^{i,j} = \int_0^t Z_\tau^{i,j} d\tau$, this is a typical problem in non-linear filtering: given a model for the dynamics (evolution in time) of x_t, for any time t, and given the past samples of the observation processes $y_\tau^{i,j}, \tau < t$, which depend on x_τ, calculate the conditional expectation $E\{x_t | y_\tau,\ \mathrm{O} \leq \tau \leq t\}$ providing the best estimate of x_t given all the observations Rewriting equation (2) in terms of the $\{y_t^{i,j}\}$'s, one finds that the dependence of the $y_t^{i,j}$ on x_t is given by

$$dy_t^{i,j} = f_{i,j}(x_t) dt + \sigma db_t^{i,j} . \qquad (3)$$

'1'bus, regardless of the motion model for x_t, our problem is indeed nonlinear because the functions $f_{i,j}$ are far from linear (indeed, they are quite localized in space).

One of the major results of non-linear filtering in the last twenty years is that such a non-linear estimation problem is in fact solvable, provided a certain second order partial differential equation can be solved. The latter is the Zakai equation (see for example [BC], [L S], [MZ], [RB], [RM], [TD]). Zakai's remarkable result has unfortunately not been widely known and used because Zakai's equation is, in typical cases, quite difficult to solve, exactly or numerically.

To preserve the flow of our exposition, we summarize the results of R. Mortensen, T. Duncan and Zakai which we will need in Appendix A, and continue here with the main topic. We just observe now that in order to use these results to deal with our problem, we still need to write down the a priori constraints on the motion of the object (i.e., the analogue in our case of equation (A 1) of the appendix). In the next section, we will look at two such motion models, and try to solve the Zakai equation in those two cases. Notice also that the components of our observation functions must be assumed differentiable in order for the formalism of stochastic filtering to apply. The observation functions are the

$$f_{i,j}(x) = 2^{\frac{j}{2}} \int_{-\infty}^{\infty} \psi_{0,0}(2^j x' - i) r(x' - x) dx' , \qquad (4)$$

and r is not necessarily continuous let alone differentiable (indeed, the object will typically have a well-defined edge, and r will then drop to O abruptly at that edge), We must therefore require that $\psi_{0,0}$ be differentiable. The Haar wavelet that we considered initially is not even continuous, but there are many other wavelets generating orthonormal bases which are (see [11)1], [11)2]).

3 The nonlinear filter

At this stage, we have formulated the problem in such a way that our observation processes do satisfy the assumptions made in the hypotheses leading to the Zakai equation. We still need to identify the state variables that must be estimated and the evolution equation governing their dynamics (i.e., the analogue of equation (A 1) for our problem). Calling X_t our vect or of state variables, heuristically it seems natural to build into our model for X_t the *continuity* of the motion of the object, as well as any *bounds on its velocity* that seem physically reasonable. A natural choice would be to take $X_t = x_t$ itself, and, lacking any a

priori knowledge about the motion of our object, to assume that the motion is Brownian. Mathematically, we would then be assuming that $x_t = \sigma' B_t$, where σ' would incorporate all our a priori assumptions about the tendency of our object to move away from x_0. While the simplicity of this assumption is appealing', it has two major drawbacks. It is not likely to reflect realistically the actual motion of many objects we would typically be interested in locating; and tracking. Perhaps more important, it will not allow us to extend these considerations to objects which do not possess the symmetry we have assumed earlier. Indeed, the two motion models we will consider in greater detail do allow the method to apply to asymmetric objects. So, instead of this natural but too simple motion model, let us consider, as a first alternative, its first integral

Still assuming for simplicity that the motion is taking place on the real line (i.e., in our original notation, with N = 1), let us assume that the motion of our object is described by a 2-dimensional process X_t, written $X_t = (x_t, v_t)$, with x representing the position of the object, v its velocity, For the dynamics, we assume that

$$\begin{aligned} \frac{dx_t}{dt} &= v_t, \\ v_t &= \sigma' B_t, \end{aligned}$$

where B_t is Brownian motion. This is the analogue of equation (A 1) which describes this first motion r-node]. It says that the velocity of our object evolves according to a Brownian motion starting from an initial velocity v_0. It is therefore natural to call this model the integral of Brownian motion. The corresponding Zakai equation can now be obtained by identifying the various terms in the general version given in the appendix. One finds that, in this case,

$$\frac{\partial \tilde{\rho}}{\partial t} = \frac{\sigma'^2}{2} \Delta_v \tilde{\rho} - v \cdot \nabla_x \tilde{\rho} - \left(\frac{1}{2\sigma^2} \sum_{i,j} f_{i,j}^2(x) + \frac{1}{\sigma^2} \sum_{i,j} y^{i,j}(t) v \cdot \nabla f_{i,j}(x) \right) \tilde{\rho}, \tag{5}$$

where Δ_v denotes the Laplacian in v (explicitly, in the one-dimensional case, $\Delta_v = \frac{\partial^2}{\partial v^2}$), and ∇_x denotes the gradient in x (again, in the one-dimensional case, $\nabla = \frac{\partial}{\partial x}$).

While the general form of the Zakai equation looked quite complicated, its specialization in 1 his case gives a simpler equation. The rigl,t-hand-side of (5) consists of three terms: a "diffusion" in v only, a "drift" in x, and a multiplication term. Still, solving such an equation exactly is quite difficult. However we can obtain an approximate solution if we "group" the terms on the right judiciously. Indeed, let us call U the function $U(t, x, v) = \frac{1}{2\sigma^2} \sum_{i,j} f_{i,j}^2(x) + \frac{1}{\sigma^2} \sum_{i,j} y_{i,j}(t) v \cdot \nabla f_{i,j}(x)$ appearing as the multiplication term, and rewrite the operators in the right-hand-side of (5) as

$$\begin{aligned} D_1 &= \frac{\sigma'}{2} \Delta_v, \\ D_2 &= -v \nabla_x - U, \end{aligned}$$

so that the Zakai equation itself can now be written as

$$\frac{\partial \tilde{\rho}}{\partial t} = D_1 \tilde{\rho} + D_2 \tilde{\rho}. \tag{6}$$

To proceed, let us make an analogy with linear systems of first order ordinary differential equations. To find a vector-valued function of time $Y(t)$ solving a system $\dot{Y} = (A_1 + A_2)Y$, where A_1, A_2 are constant matrices, one simply computes $Y(t) = \exp(t(A_1 + A_2)) \cdot Y(0)$, where $Y(0)$ is a prespecified initial condition. Moreover, if the sum $A_1 + A_2$ on

the ri.gilt-hanct-side is difficult to exponentiate, one can use the formula $\exp(t(A_1+A_2)) = \lim_{n\to\infty}(\exp(\frac{t}{n}A_1)\exp(\frac{t}{n}A_2))^n$, which is quite helpful if the exponential of each summand is known. In fact, one does not need to let n tend all the way to ∞ in this last formula if one is only interested in an approximate answer. Indeed, with n = 1, one can still say that if t is an arbitrary point in time, and if >'(f) is known, then for δ sufficiently small the solution at $t+\delta$ can be approximated by $Y(t+\delta) \approx \exp(\delta A_1)\exp(\delta A_2)Y(t)$. Moreover, it turns out that the $\approx$ sign in this last formula can in fact be replaced by an exact equality if the matrices A_1 and A_2 commute.

in our case, these considerations have a natural generalization. If we choose a time-increment δ that is small enough that the function $E(t,x,v)$ can be considered almost independent of time on the interval $[t, t+\delta]$, one can then write down an approximate solution to (6) in the form

$$\tilde{\rho}(t+\delta, x, v) \approx \exp(\delta D_1)\cdot\exp(\delta D_2)\cdot\tilde{\rho}(t,x,v), \tag{7}$$

where the operators on the right are defined by

$$(\exp(\delta D_1)\cdot p)(x,v) = \int p(x,v')\frac{e^{-\frac{(v-v')}{(2\sigma'^2\delta)}}}{\sqrt{2\pi\sigma'^2\delta}}dv', \tag{8}$$

$$(\exp(\delta D_2)\cdot p)(x,v) = p(x-\delta v, v)e^{-\int_0^\delta E(x-\tau v,v)d\tau}. \tag{9}$$

for p a function of (x,v).

Equation (7) gives a recursive procedure for finding an approximate solution to the Zakai equation in this case, i.e. when the motion model is the integral of Brownian motion: starting with the initial density $\tilde{\rho}_0$, "advance it" in time by small time-increments according to (7). Heuristically, at each iteration, the first step is to account for the randomness in the motion that could nave affected the position of the object between time t and $t+\delta$; the second step is to incorporate the information contained in the observations $f_{i,j}$. That is what equations (8) and (9) accomplish. A more general version of this idea was originally proposed and extensively developed in [HW]. In our case, as we will see in the next section, this simple approximation will actually give the exact solution in the simplest case.

For the second motion model, we adapt an idea from V. Benes's considerations [VB]. Let us refer to it as the Poisson model. We will still need our two variables x_t and v_t, subject to the assumptions $\frac{dx_t}{dt} = v_t$, and that the velocity process v_t takes independent values, with density g, which are constant between jumps of a Poisson process of rate λ. This means that our object moves in straight line segments, changing its course at a sequence of times $\{t_1, t_2, t_3, \ldots\}$ such that the quantities $t_1, t_2 - t_1, t_3 - /2, \ldots$ are distributed according to a Poisson distribution of mean $1/\lambda$, and such that the new velocities at each course change are distributed according to a prespecified density function g. The corresponding Zakai equation is

$$\frac{\partial\tilde{\rho}}{\partial t} = -v\cdot\nabla_x\tilde{\rho} + \lambda g(v)\int\tilde{\rho}(t,x,v')dv' - \left(\frac{1}{2\sigma^2}\sum_{i,j}f_{i,j}^2(x) + \frac{1}{\sigma^2}\sum_{i,j}y^{i,j}(t)v\cdot\nabla f_{i,j}(x) + \lambda\right)\tilde{\rho} \tag{10}$$

To solve this equation, let us write $\tilde{\rho} = \sum_{n\geq 0}\tilde{\rho}_n$, and require that each $\tilde{\rho}_n$ satisfy

$$\frac{\partial\tilde{\rho}_n}{\partial t} = -v\cdot\nabla_x\tilde{\rho}_n + \lambda g(v)\int\tilde{\rho}_{n-1}(t,x,v')dv' - \left(\frac{1}{2\sigma^2}\sum_{i,j}f_{i,j}^2(x) + \frac{1}{\sigma^2}\sum_{i,j}y^{i,j}(t)v\cdot\nabla f_{i,j}(x) + \lambda\right)\tilde{\rho}_n, \tag{11}$$

where $\tilde{\rho}_{-1}$ is assumed to be identically zero. Indeed, the requirement that $\tilde{\rho}$ satisfy the Zakai equation (10) is exactly equivalent to the

requirement that each $\tilde{\rho}_n$ satisfy equation (11). The introduction of the index n makes it possible to solve this system of inhomogeneous first order equations by induction, using the method of characteristics. in fact, if we set

$$F(t,x)=\frac{1}{2\sigma^2}[\sum_{i,j} f_{i,j}(x)^2+2y^{i,j}(t)v\cdot\nabla f_{i,j}(x)+\lambda],$$

the equation for $\tilde{\rho}_0$ is solved by

$$\tilde{\rho}_0(t,x,v)=\tilde{\rho}(0,x-tv,v)e^{-\int_0^t F(s,x-(t-s)v)ds},$$

where the initial density $\tilde{\rho}(0,x,v)$ is assumed known. Practically, a slightly more general version of this equation, relating $\tilde{\rho}$ at time t to $\tilde{\rho}$ at any previous instant in time, would be useful, This can be obtained by writing the solution for $\tilde{\rho}_n$, n $\geq$ O as

$$\tilde{\rho}_n(t,x,v)=\tilde{\rho}_n(t_0,x-(t-t_0)v,v)e^{-\int_{t_0}^t F(s,x-(t-s)v)ds}+\lambda g(v)\int_{t_0}^t[\int\tilde{\rho}_{n-1}(s,x-(t-s)v,v')dv']e^{-\int_s^t F(\tau,x-(t-\tau)v)d\tau}ds \quad (12)$$

keeping in mind that $\tilde{\rho}_{-1}$ is zero. That these formulas for $\tilde{\rho}_n$ do give solutions to the prescribed equations for $\tilde{\rho}_n$ can be verified directly. One can now reconstitute $\tilde{\rho}$ by adding all these equations together. Remarkably, the equat ion one then gets is exactly (12) without the index n ! The left-hand-sides add up to $\tilde{\rho}$ at time t, and the right-hand-sides involve the values of $\tilde{\rho}$ at earlier times.

in this case, i.e. when the motion model is the Poisson model, one can thus again comput e ρ by "recursion" on t, this time *exactly*. As before, at each new time t, the iteration proceeds in two steps: first the motion randomness is accounted for (witness the terms "$\tilde{\rho}(t_0,x-(t-t_0)v,v)$" and "$\tilde{\rho}(s,x-(t-s)v',v)$" in (1 2), then the information in the new observations is incorporated (the exponential terms in (1 2)). The form of the solution in this case highlights quite clearly the tremendous simplification one enjoys thanks to the orthogonality of the wavelet basis. This property is not Crucial to the derivation of the equation governing the corresponding conditional density. However, if it is not made, one would end up with a huge number of "cross-t terms" in the exponents appearing in the formula for $\tilde{\rho}_n$ (and thus for $\tilde{\rho}$) above, as well as in the expression relating ρ and $\tilde{\rho}$. These cross-terms would be due to the correlation between the noise in the different observations. They would practically manifest themselves as double sum (over $(i,j),(i',j')$) instead of the single sums we now end up with, and would then make any implementation of the formulas prohibitively time consuming. The orthogonality assumption has thus allowed us to replace impossibly "dense" matrix multiplications of the form $\sum_{i,j,i',j'} A^{i,j}_{i',j'} f_{i,j}$ (where A would have represented essentially the noise correlation mat rix) by manageable "diagonal" sums of the form $\sum_{i,j}$ such as we now have,

Before going on to the examples, it is important to notice that we only needed the orthogonality of the $\{\psi_{i,j}\}$ and not their completeness. The results we nave obtained would thus remain valid if only a finite subset of them is used, which is indeed what must happen in practice, While there is in general no reason to select a priori such a finite subset, there is nothing in the derivation of the formulas preventing us from doing so. Indeed, $\{\psi_{i,j}\}$ could be any orthonormal set of continuously differentiable square-integrable functions.

Finally, let us reconsider one other simplifying assumption which can be removed. That is the assumption about the symmetry of r. Consider an object moving in the x^1x^2-plane. If we merely assume that the object is rigid (with a possibly complicated asymmetric shape), the pattern it makes on an image will depend on its velocity vector as well as on its coordinates. in fact, if $r(x^1,x^2)$ denotes the pattern it makes when it is placed at the origin "lacing" in the direction of the positive

x^1-axis, its presence at (x_t^1, x_t^2) at time t, with velocity (U: , v_t^2), will be represented by the intensity function

$$R(x^1 - x_t^1, x^2 - x_t^2; v_t^1, v_t^2) =$$
$$r\left(\frac{v_t^1(x^1-x_t^1)+v_t^2(x^2-x_t^2)}{\sqrt{(v_t^1)^2+(v_t^2)^2}}, \frac{v_t^2(x^1-x_t^1)+v_t^1(x^2-x_t^2)}{\sqrt{(v_t^1)^2+(v_t^2)^2}}\right),$$

since the matrix expressing the rotation of the positive x^1-axis to the positive (v_t^1, v_t^2)-axis is exactly

$$\frac{1}{\sqrt{(v_t^1)^2 + (v_t^2)^2}} \begin{pmatrix} v_t^1 & -v_t^2 \\ v_t^2 & v_t^1 \end{pmatrix}$$

'1'bus, if ψ is any member of the orthonormal set we have chosen to use to make our observations, one can then compute the associated observation function f (see equation (4)) in the form

$$f(x_t^1, x_t^2, v_t^1, v_t^2) =$$
$$\iint \psi(x^1, x^2) R(x^1 - x_t^1, x^2 - x_t^2; v_t^1, v_t^2) dx^1 dx^2$$

4 Examples

The first example illustrates the case where the Poisson and $\int$-Brownian models are exactly equivalent, namely when λ = O, u' :: O. One can verify directly that the formulas for the conditional density in the two cases are then identical and give the exact density function. In that case, we are tracking a marker moving on a fixed course with constant velocity. For the example, we chose the "observations" to be 1-dimensional images of a scene consisting of a marker moving within the interval [0, $128\Delta x$] on the real line. The "marker" is given by the function $r(x) = 1$ if $|x| < 3\Delta x/2$, 0 otherwise, translated to whatever position its "center" occupies. Snapshots of the scene at 25 instants of time Δt apart were synthesized. The velocity v_t was kept constant; for simplicity, we chose $v =$ -- A x/ A t. The funct ion $\psi_{0,0} = \psi$ was chosen to be the *scaling function* for the "cubic spline wavelet" of Y. Meyer (pictured in [SM]), rather than the wavelet itself. It is continuously differentiable, and equal to a cubic polynomial on every interval $[n\Delta x, (n+1)\Delta x]$. Its translates do form an orthonormal family, as required. We used the 128 observations $Z^{i,0}$, $(1 \le i \le 128)$, obtained using the functions $\psi(x - i\Delta x)$. The size and resolution of the images wc synthesize and the scale of t he motion justify this choice of the function ψ and the corresponding finite set of observations. The individual scenes were synthesized by starting the marker at $x_0 =$ 96A.?', and moving it at speed v from there on, adding "noise" t hat is piecewise constant on every interval $[n$ A $x, (n$ -t 1) A $x]$, with peak values chosen independently from a O-mean normal distribut ion. The σ in the equation for the observation process then turns out to be the root-nlean-square height of these peaks. Finally, we assumed that the initial density is uniform over the interval [0, 128A2].

Figure 1 shows one su ch set of "sn apshots" (time in creases upward), with σ = .1. Figure 2 shows another one, with $\sigma = 1$. The (normalized, discrete version of the) density function $\rho_{k\Delta}(x)$, and the optimal filter $\hat{x}_{k\Delta} = \int x \rho_{k\Delta}(x) dx$ can be computed using the formula for the solution of the Zakai equ at ion. The values of $\hat{x}_{k\Delta}$ are shown to the right of the corresponding "snapshot" in Figures 1 and 2. The accuracy of the result is hardly surprising. One could have obtained the location of the marker, apparently more simply, by "stacking" the appropriately laterally shifted signals on top of one another (i.e. adding the suitably translat ed signals) to increase the "signal-to-noise" t atio and thus extract the position of the marker out of the noise. In effect, that is exactly what our algorithm does in this case, only better so in the sense that it is optimal

(in particular, it draws the appropriate conclusions at the earliest opportunity)

The "shift-and-sum" approach is not applicable in the case of a "maneuvering" object, however. The second example illustrates the formulas we derived in the previous section in this case. in this example, we used the Poisson motion model with $1/\lambda = 5\Delta t$ to generate a path for the marker, with a two-point density function for the velocity, namely $g(\exists A\ x/\Delta t) = 1/2$. *I'bus,* the marker maintains the same speed through its direction changes. Figure 3*a* shows the snapshots synthesized with $\sigma = 1/2$, with the actual location of the center of the marker to the right of the corresponding image. Figure *3b shows* the graphs of the corresponding normalized densities $\rho_{k\Delta t}$. The figure to the right of each graph is its expected value, i.e. the value of $\hat{x}_{k\Delta t}$. Note that, in this case, the evolution of the position of the marker can be visually estimated by a cunning observer: indeed, by tilting the page in such a way that the grazing angle of the rays from the page to the eye of the astute observer is very shallow, she (or he) can guess fairly accurately the position of the marker on every frame. The general form of the algorithm of the previous section allows one in effect to generalize this "tilting" approach to the 2-dimensional case (where it cannot be carried out visually), again in an optimal way. Perhaps more remarkably, the algorithm tracks a marker even when the '[tilting" approach in the l-dimensional case eludes the most astute observer. Figure 4a shows snapshots synthesized as in Figure 3*a* but with $\sigma = 1$. An honest observer would find it difficult to spot the position of the marker visually with any precision (of course, a more accurate assessment would have to take into account the recognition differential of the typical human observer). Figure 4*b* shows the graphs of the densities $\rho_{k\Delta t}$. Their accuracy in finding the marker (albeit after an initial time of uncertainty of duration $\approx 7\Delta t$) is very encouraging.

Let us now look at an example of the two-dimensional situation. In this instance, the filter was designed assuming a Poisson motion model to find the simulated track of a small disk moving over the water on a Synthetic Aperture Radar image of the Pacific ocean off Mission Beach. The disk was "synthetically" added to the original image. The SAR resolution was about 10 meters, and we assumed the object was a disk with a 5-meter diameter. A piecewise linear track was generated using a random number generator to choose the straight- course durations, assuming that the new velocity at every course change is chosen from a set of 23 possibilities, with probability as described in Figure 5. One can immediately anticipate that the resolution of the SAR images compared to the size and motion parameters of the object should create a substantial ambiguity in the localization of the object: indeed the object may in reality move according to the model used without having a greatly perceivable effect on the pixel intensities. Nevertheless, one can still ask within which pixel the object is at any point in time. To answer that question using our method, the motion model chosen to design the filter was slightly different from the one which generated the actual track: vie used the filter derived for the Poisson model, assuming that the marker was likely to change course on average once every time step, with its new velocity one of the nine possibilities associated to each of the nine pixels adjacent to the one it "occupies". That is appropriate when the conditional density is in the first place approximated by an admittedly inexact but adequate matrix whose entries are meant 10 represent the nearly constant value of the density over each pixel. The result is a quite a-curate localization algorithm (see figs. 6 and 7). For the first test, the additive white gaussian noise was generated so that its variance would be that of the sample variance of

the original pixel intensities over the left half of the picture (roughly, the ocean part of the scene). The intensity of the marker (per unit area) was chosen to equal that of the brightest spot, on that left portion. Due to the small area of the disk, this translates to a signal-to-noise ratio of 11 .7dB at best (when the marker falls entirely within a pixel), but it could be as low as – .25dB (if the marker overlays four adjacent pixels). The track origin was chosen so that this least advantageous case occurs at the first frame. The intial density was assumed uniform over the upper left quadrant of the image. Figure G snows the successively synthesized noisy images, together with a gray-intensity plot of the updated conditional density function. The second test. was performed with the intensity of the marker reduced to achieve a signal-to-noise ratio between 7dB (in the best case) and –5dB (in the worst case, which, again, occured on the first frame). The plots of the density function in that case (fig. 7) show that in spite of some initial uncertainty, the localization is quite good after nine frames. These results confirm that one can localize the position of the object, at least to within the resolution of the original images, in the presence of substantial noise and given that the object cross-section makes it quite undetectable on any single image.

These examples also show that the algorithm is not very sensitive to changes in the form or intensity of the object. Indeed, we specifically chose an object size smaller than the dimensions of a pixel, and allowed the object to move in such a way that its intensity would contribute to one, two or even four adjacent pixels, thus making the SNR fluctuate by as much as 12dB. This robustness is encouraging. The computation efficiency of the algorithm is less impressive. The last example was implemented on a SUN Spare 2 workstation, and required several minutes of processing time. It is therefore clear that the method in its present form is interesting only in those cases where the SNR is so low that the object is not detectable on a single image, i.e. in those cases where tracking is necessary *in order to* detect and localize.

References

[BC] R.W. Brockett and J .M.C. Clark, *The geometry of the conditional density equation*, in Analysis and optimization of stochastic systems, O. Jacobs et al. eds., 1980, pp. 299-309.

[HW] W.E. Hopkins and W.S. Wong, *Lie-Trotter product formulas for nonlinear filtering*, Stochastics, 17 (1985), pp. 313-337.

[HN] H.H. Nagel, *Analysis techniques for image sequences*, Proc. 4th Intl. Conf. Pattern Recognition, Nov. 1978, pp. 186-211.

[D1] 1. Daubechies, The *wavelet transform, time-frequency localisation, and signal analysis*, I.E.E.E. Trans. Information Theory, 36 (1990), pp. 961-1005.

[D2] I. Daubechies, *Orthonormal basis of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909-996.

[LS] R.S. Lipster and A. N. Shiryayev, *Statistics of random processes*, Springer Verlag, 1977.

[LY] G.R. Letgers, Jr., and T.Y. Young, *A mathematical model for computer image tracking*, I.E.E.E. Trans. P.A.M.I., 4 (1982), pp. 583-594.

[MZ] M . Zakai, *On the optimal filtering of diffusion processes*, Z. Wahrsch. verw. Gel).)ll (1969), pp. .230-243.

[RB] R.W. Brockett, *Nonlinear systems and nonlinear estimation theory*, in Stochastic Systems: The mathematics of filtering and identification and applications, M. Hazewinkel et al. eds., 1981, pp. 441-477.

[RM] R.E. Mortensen, *Optimal control of continuous time stochastic systems*, Ph. D. thesis, (University of California at Berkeley, 1966.

[SM] S. Mallat, *A compact multiresolution representation: the wavelet model*, in Proc. IEEE Wkshp. Comp. Vision, (1987), pp. 2-7.

[TD] T.E. Duncan, *Probability densities for diffusion processes with applications 10 nonlinear filtering theory*, Ph.D. thesis, Stanford, 1967.

[VB] V.E. Benes, *Nonlinear filtering in sonar detection and estimation*, preprint

Appendix A

Here is a simplified summary of those results of R. Mortensen, T. Duncan and Zakai which we will need (for a complete account, see [MZ], [RM],[TD], or [1,S]). Assume that X_t is a stochastic process in $\mathbf{R}^n$ satisfying the stochastic equation

$$dX_t = h(X_t)dt + g(X_t)dB_t\,, \qquad \text{(A1)}$$

where $h : \mathbf{R}^n \to \mathbf{R}^n$ and $g : \mathbf{R}^n \to \{n \times m \text{ matrices}\}$ are twit.c-diflerentiahle functions, and B_t is Brownian motion in R'", Assume we dispose of observations Y_t (a stochastic process in $\mathbf{R}^k$) which obey

$$dY_t = f(x_t)dt + \sigma db_t\,, \qquad \text{(A2)}$$

where σ is a real constant, j : $\mathbf{R}^n \to \mathbf{R}^k$ is a differentiable function, and b_t is Brownian motion in $\mathbf{R}^k$. Assume further that $B_t, b_{t'}$ and x_0 are independent for all t different from t'. Given an initial density ρ_0 for the states of X_0, the conditional density $\rho(t, x$ given Y_s for s $\leq$ t) for X_t, given all Y_s up to time t, must satisfy the (Zakai) equation

$$\frac{\partial \tilde{\rho}}{\partial t} = D\tilde{\rho} - \frac{1}{\sigma^2} Y_t \cdot D'\tilde{\rho} + \frac{1}{\sigma^2}\left(Y_t \cdot D'' Y_t\right)\tilde{\rho}\,,$$

where ρ and $\tilde{\rho}$ are related by $\rho(t,x) = a(t)e^{Y_t \cdot f(x)/\sigma^2}\tilde{\rho}(t,x)$ in which the function o takes that value making $\int_J \rho(t,x)dx = 1$ for all t, and where the operator D is given by

$$D\tilde{\rho} = \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\left((gg^T)_{ij}\tilde{\rho}\right) - h \cdot \nabla\tilde{\rho} - (\text{div } h)\tilde{\rho} - \frac{1}{2\sigma^2}|f|^2\tilde{\rho}\,,$$

(here g^T denotes the transpose of the matrix g, and div h denotes the divergence of h), the i^{th} component of the vector operator D' is given by

$$(D'\tilde{\rho})_i = [f_i, D]\tilde{\rho}\,,$$

(where the term on the right refers to the commutator of the two operators "multiplication by f_i" and D), and the matrix operator D'' is given by

$$(D''\tilde{\rho})_{ij} = \frac{1}{2}[f_i, [f_j, D]]\tilde{\rho}$$

Once ρ has been computed hy solving the Zakai equation with initial condition $\rho|_{t=0} = \rho_0$, the required conditional expectation $E\{x_t |$ y,, O $\leq \tau \leq t\}$ can be computed by performing the integral $E\{x_t | y_\tau$, O $\leq \tau \leq$ $t\}$ = $\int x\rho(t,x)dx$. (Note that if f, g and h are linear, i.e when the filtering problem is linear, Kalman snowed that the conditional expectation can t hen be calculated directly by solving a coupled system consisting of a linear equation in n variables driven by the observations,

coupled to a Riccati equation in $n(n+1)/2$ variables involving the second moment only. That is infinitely more manageable than having to find the function ρ of n + 1 variables solving the Zakai partial differential equation).

Figure Captions

Figure 1

'1'wenty-five l-dimensional consecutive "snapshots", with time increasing upward, The noise was synthesized to get $\sigma = .1$; the figures on the right are the values of the optimal filter $\hat{x}$; in this case, they equal the (horizontal) coordinate of the position center of the marker exactly.

Figure 2

Same as in Figure 1, except that the noise here was synthesized to get $\sigma = 1$; note that the filter output remains within $3\%\Delta x$ of the exact position as of time $t = 7$At in spite of the fact that x_0 was assumed uniformly distributed over the interval $[0, 128\Delta x]$.

Figure 3a

Snapshots of a marker moving in noise generted to obtain $\sigma = .5$; the exact positions of the center of the marker were.gcncratcd using the Poisson model with a rate of $.2/\Delta t$, and are given by the figures to the right fo each "snapshots".

Figure 3b

The graphs of $\rho_{j\Delta t}$ calculated from the observations made from Figure 3a, with j increasing from Δt to $25\Delta t$ from top to bottom. The figures on the right are the values of the optimal filter $\hat{x} = \int x\rho_{j\Delta t}(x)dx$.

Figure 4a

Same as Figure 3a, except that the noise here gives $\sigma = 1$.

Figure 4b

Same as Figure 3b, using the observations made from Figure 4a.

Figure 5a

Table showing the probability density function of the new velocity V at each course change, where $V = |V|(\cos\theta, \sin\theta)$; $|V|$ is in units of At'/At, where Δx is the width of a pixel, and O in degrees. The difference between the sum of all the entries and 1 is the probability that $V = 0$.

Figure 5b

Sample track used in simulations. The average rate of velocity change was chosen to be $.4/\Delta t$.

Figure 5c

Synthesized noisy image of that 128 x128-pixel portion of the original image which wc used in the simulations.

Figure 6

Consecut ive SAR images with noise and evolving marker, and the corresponding intensity plots of the updated conditional density function describing the position of the marker. The dark spot in the upper left portion of the first intensity plot is at the correct marker position at the initial time. The intensity of the marker was chosen to achieve a "signal-to-noise" ration between $-.25$dB and 11.7dB (depending on whether the marker ison a pixel or between pixels).

Figure 7

Intensity plots of the updated conditional density function describing the position of the marker, corresponding; to consecutive SAR images with noise and evolving marker, as in Figure 6, except that the "signs]-to-noise" ratio is between −5dB and 7dB.

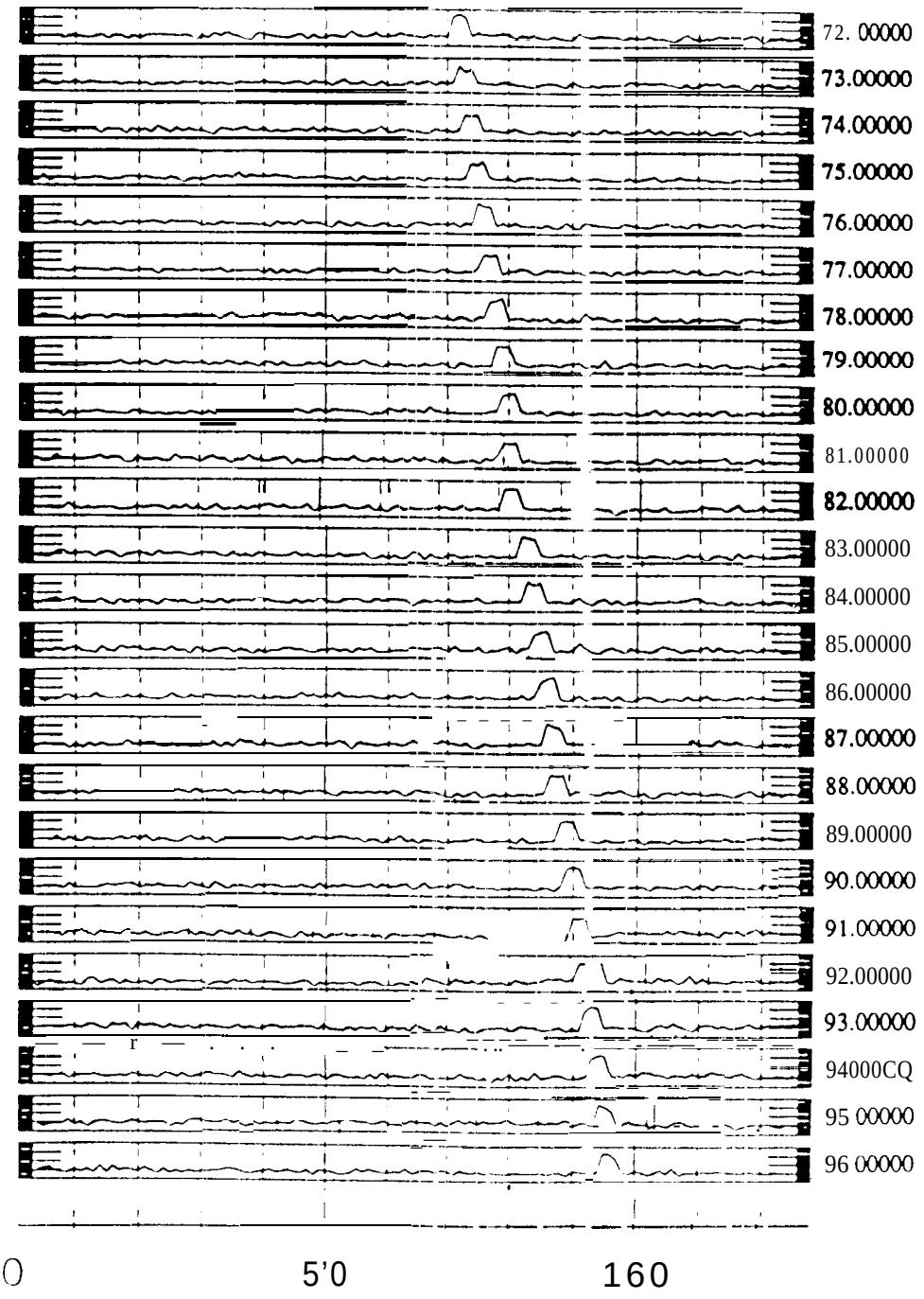

72. 00000
73.00000
74.00000
75.00000
76.00000
77.00000
78.00000
79.00000
80.00000
81.00000
82.00000
83.00000
84.00000
85.00000
86.00000
87.00000
88.00000
89.00000
90.00000
91.00000
92.00000
93.00000
94000CQ
95 00000
96 00000
0
5'0
160

Figure 1

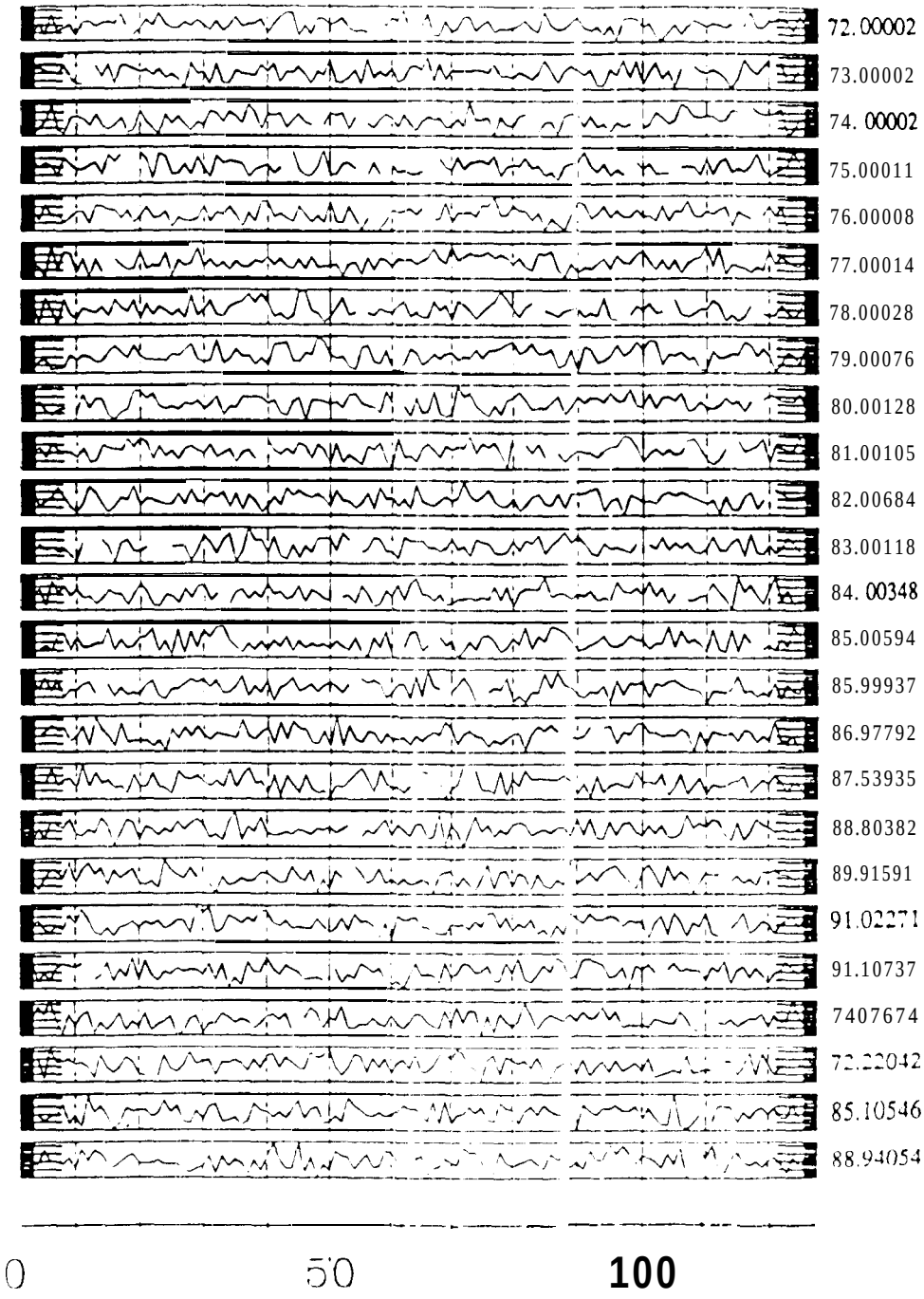

72.00002
73.00002
74. 00002
75.00011
76.00008
77.00014
78.00028
79.00076
80.00128
81.00105
82.00684
83.00118
84. 00348
85.00594
85.99937
86.97792
87.53935
88.80382
89.91591
91.02271
91.10737
7407674
72.22042
85.10546
88.94054
0
50
100

Figure 2

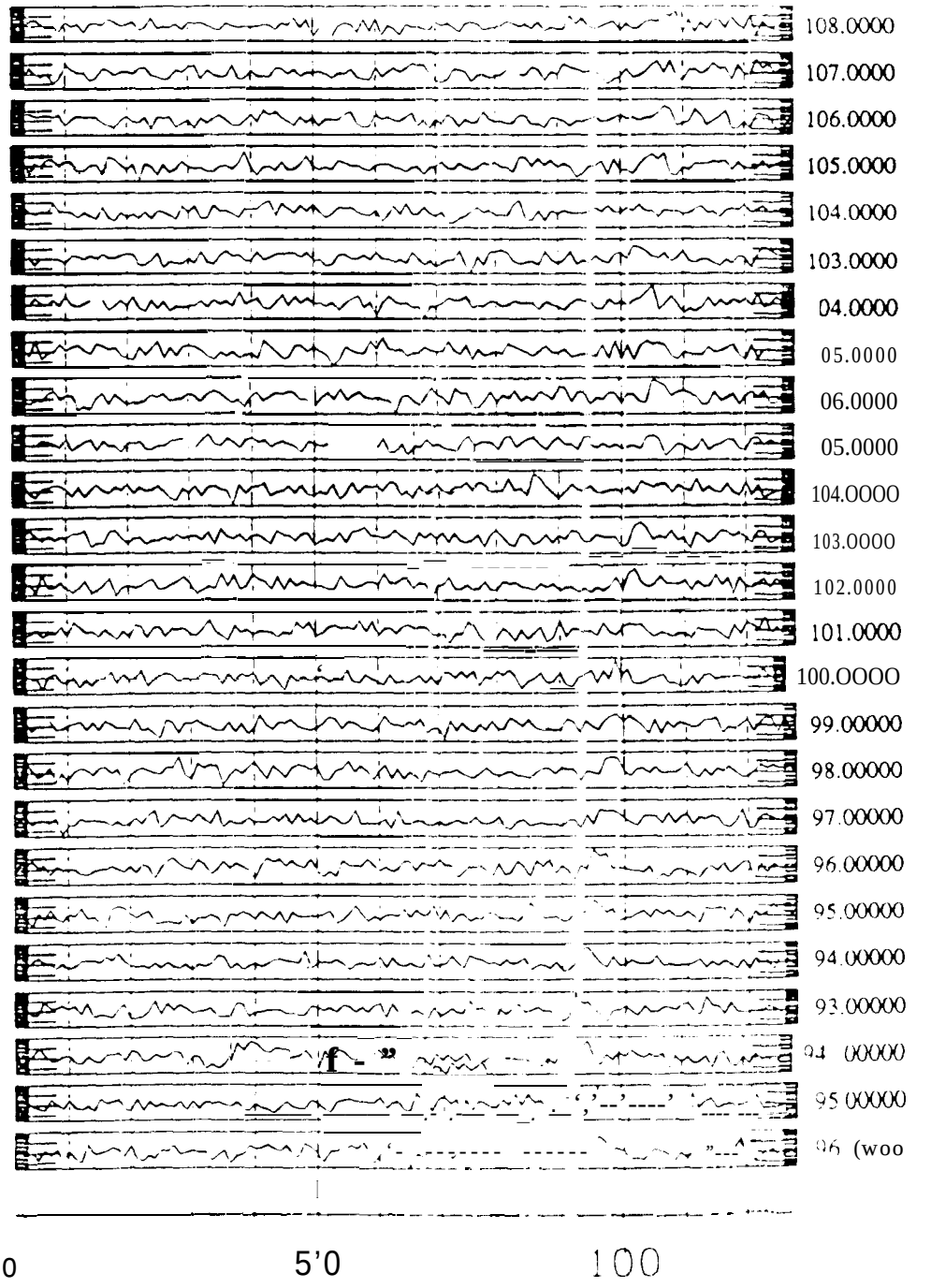

108.0000
107.0000
106.0000
105.0000
104.0000
103.0000
04.0000
05.0000
06.0000
05.0000
104.0000
103.0000
102.0000
101.0000
100.0000
99.00000
98.00000
97.00000
96.00000
95.00000
94.00000
93.00000
95.00000
0
5'0
100

Figure 3a

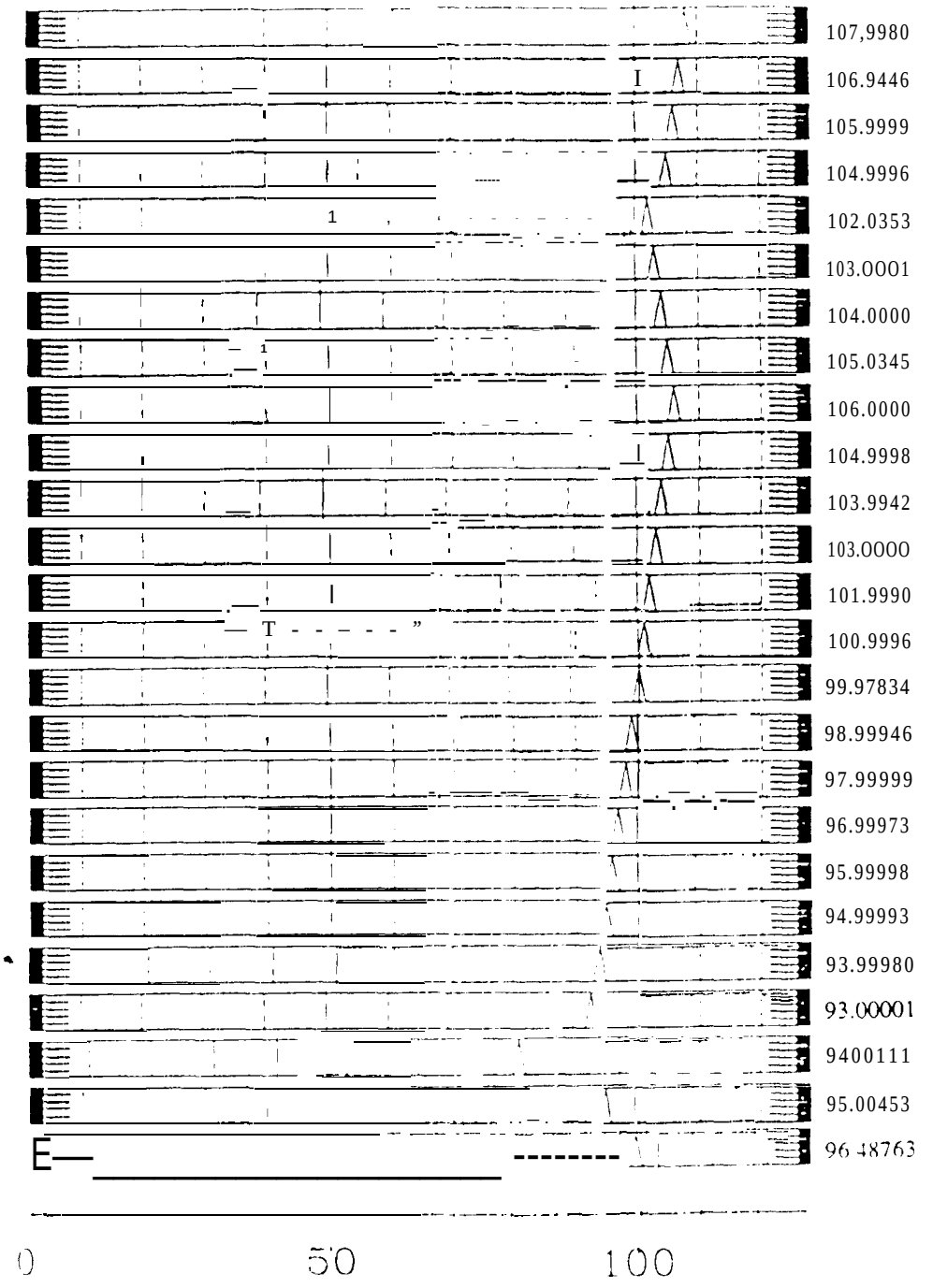

107,9980
106.9446
105.9999
104.9996
102.0353
103.0001
104.0000
105.0345
106.0000
104.9998
103.9942
103.0000
101.9990
100.9996
99.97834
98.99946
97.99999
96.99973
95.99998
94.99993
93.99980
93.00001
9400111
95.00453
96.48763
0
50
100

Figure 3b

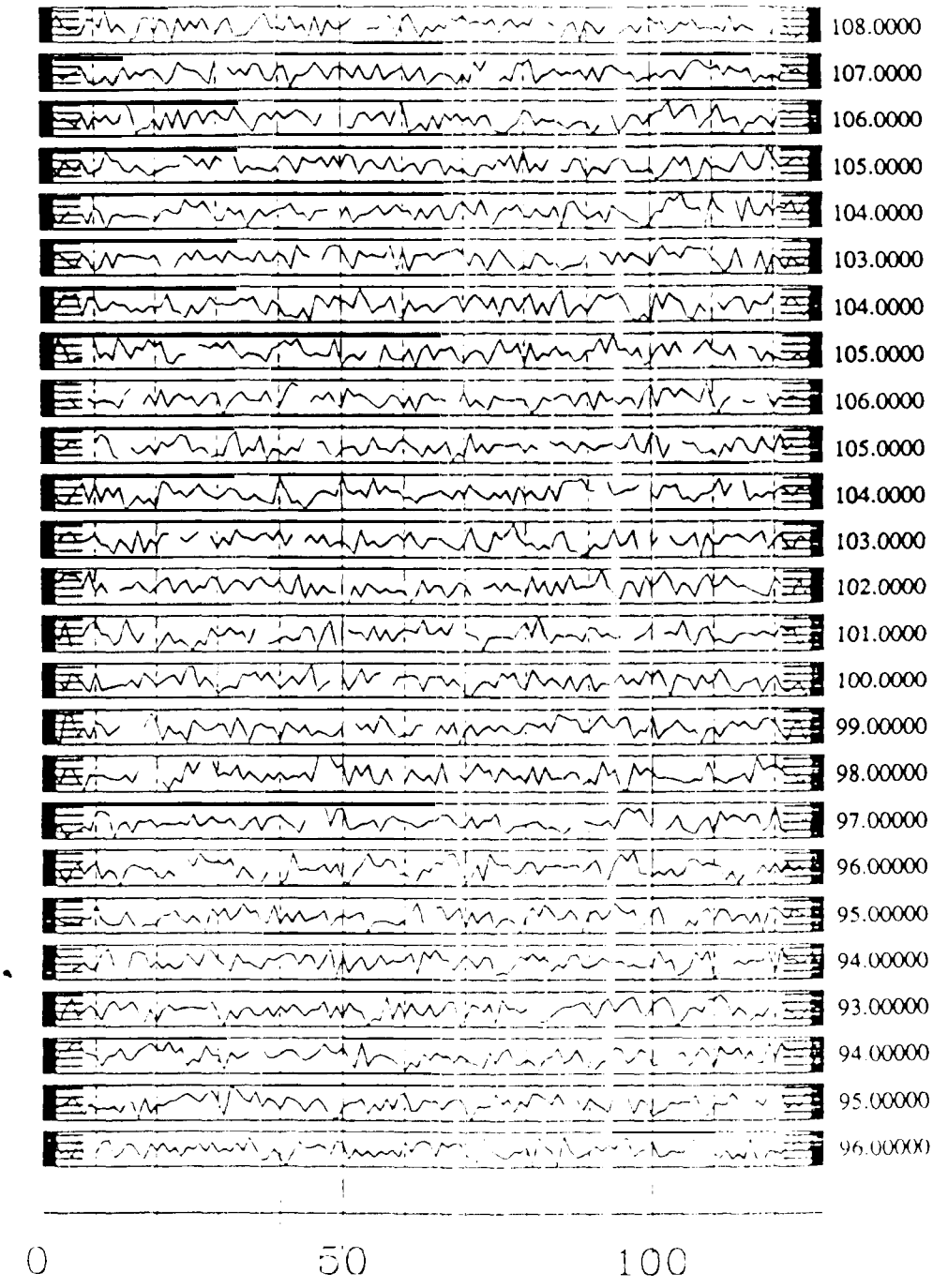

108.0000
107.0000
106.0000
105.0000
104.0000
103.0000
104.0000
105.0000
106.0000
105.0000
104.0000
103.0000
102.0000
101.0000
100.0000
99.00000
98.00000
97.00000
96.00000
95.00000
94.00000
93.00000
94.00000
95.00000
96.00000
0
50
100

Figure 4a

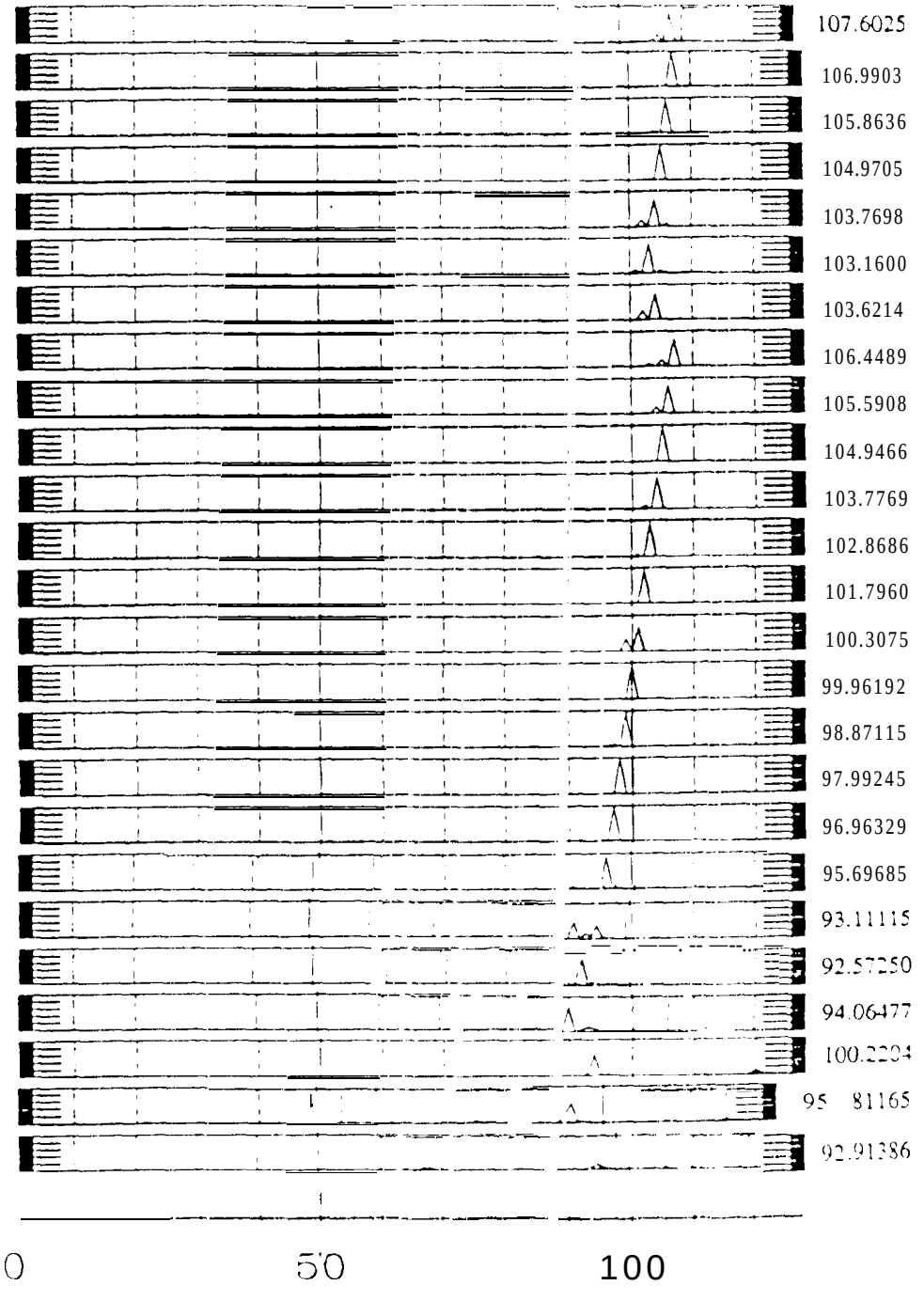

107.6025
106.9903
105.8636
104.9705
103.7698
103.1600
103.6214
106.4489
105.5908
104.9466
103.7769
102.8686
101.7960
100.3075
99.96192
98.87115
97.99245
96.96329
95.69685
93.11115
92.57250
94.06477
100.2204
95 81165
92.91386
0
50
100

Figure 4b

Figure 6

Consecutive SAR images with noise and evolving marker, and the corresponding intensity plots of the updated conditional density function describing the position of the marker. The dark spot in the upper left portion of the first intensity plot is at the correct marker position at the initial time. The intensity of the marker was chosen to achieve a "signal-to-noise" ratio between -.25dB and 11.7dB (depending on whether the marker is on a pixel or between pixels).

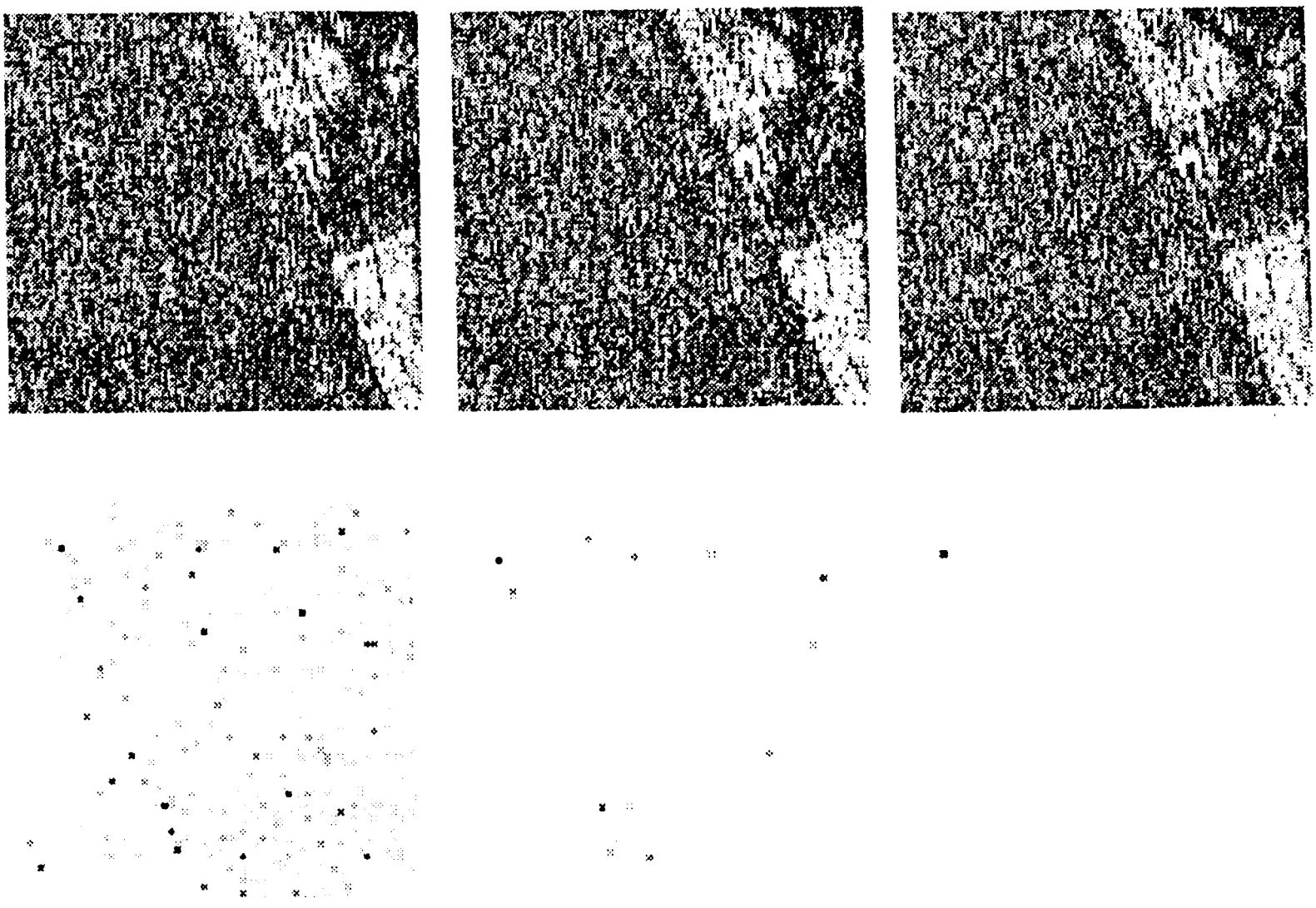
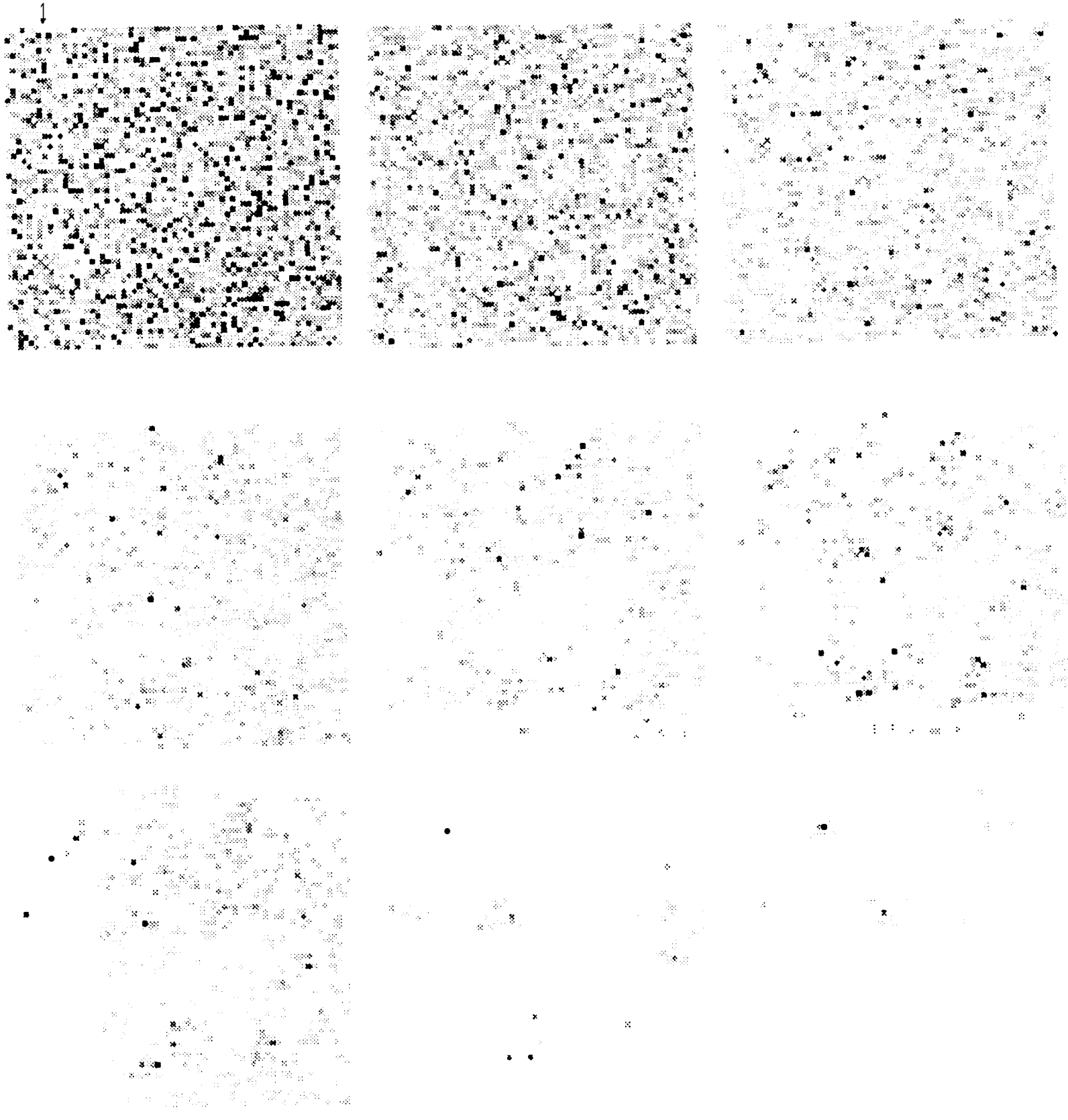


Figure 7

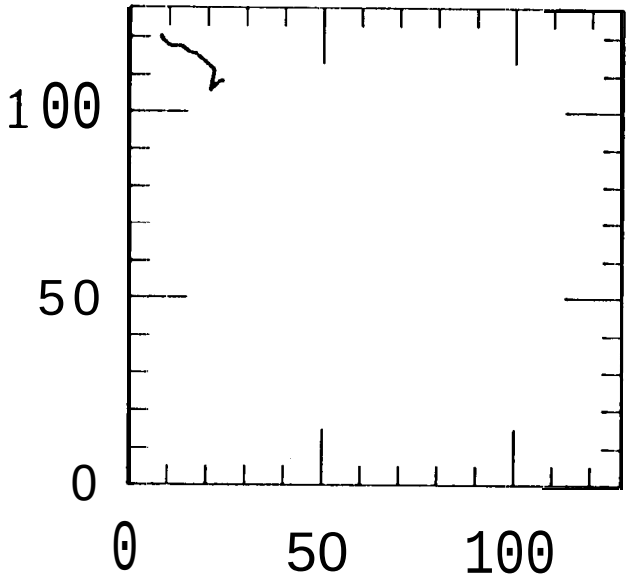


Figures 5 a, b, c

θ	-160	-140	-120	-100	-80	-60	-40	-20	0	20	40
$\|V\|=1/2$	2/87	8/261	10/261	4/87	14/261	16/261	14/261	4/87	10/261	8/261	6/261
$\|V\|=1$	2/87	8/261	10/261	4/87	14/261	**16/261**	**14/261**	4/87	10/261	8/261	6/261

Table showing the probability density function of the new velocity Vat each course change, where $V = |V|(\cos\theta, \sin\theta)$; $|V|$ is in units of $\Delta x/\Delta t$, where Δx is the width of a pixel, and θ in degrees. The difference between the sum of all the entries and 1 is the probability that $V = O$.

Sample track used in the simulations. The average rate of velocity change was chosen to be .4/At,



Synthesized noisy image of that 128x128-pixel portion of the original image which we used in the simulations.